Homework_week_2

Christophe

September 5, 2015

Homework problems: 2.6, 2.8, 2.20, 2.30, 2.38, 2.44

2.6 Dice rolls. If you roll a pair of fair dice, what is the probability of

```
x <- c(1:6)
y <- c(1:6)
M <- x[1] + y
M <- rbind(M, x[2] + y)
M <- rbind(M, x[3] + y)
M <- rbind(M, x[4] + y)
M <- rbind(M, x[5] + y)
M <- rbind(M, x[6] + y)
M <- matrix(M, ncol = 6)
M
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    2    3    4    5    6    7
## [2,]    3    4    5    6    7    8
## [3,]    4    5    6    7    8    9
## [4,]    5    6    7    8    9   10
## [5,]    6    7    8    9   10   11
## [6,]    7    8    9   10   11   12
```

(a) getting a sum of 1?

```
paste("The possibilites of rolling a 1 with two dies is",
      sum(M == 1), "out of", length(M), "possible combinations",
      "or a probability of", round(sum(M == 1)/length(M),2))
```

[1] "The possibilites of rolling a 1 with two dies is 0 out of 36 possible combinations or a probability of 0"

(b) getting a sum of 5?

```
paste("The possibilites of rolling a 5 with two dies is", sum(M == 5),
      "out of", length(M), "possible combinations",
      "or a probability of", round(sum(M == 5)/length(M),2))
```

[1] "The possibilites of rolling a 5 with two dies is 4 out of 36 possible combinations or a probability of 0.11"

(c) getting a sum of 12?

```
paste("The possibilites of rolling a 12 with two dies is", sum(M == 12),
      "out of", length(M), "possible combinations",
      "or a probability of", round(sum(M == 12)/length(M),2))
```

[1] "The possibilites of rolling a 12 with two dies is 1 out of 36 possible combinations or a probability of 0.03"

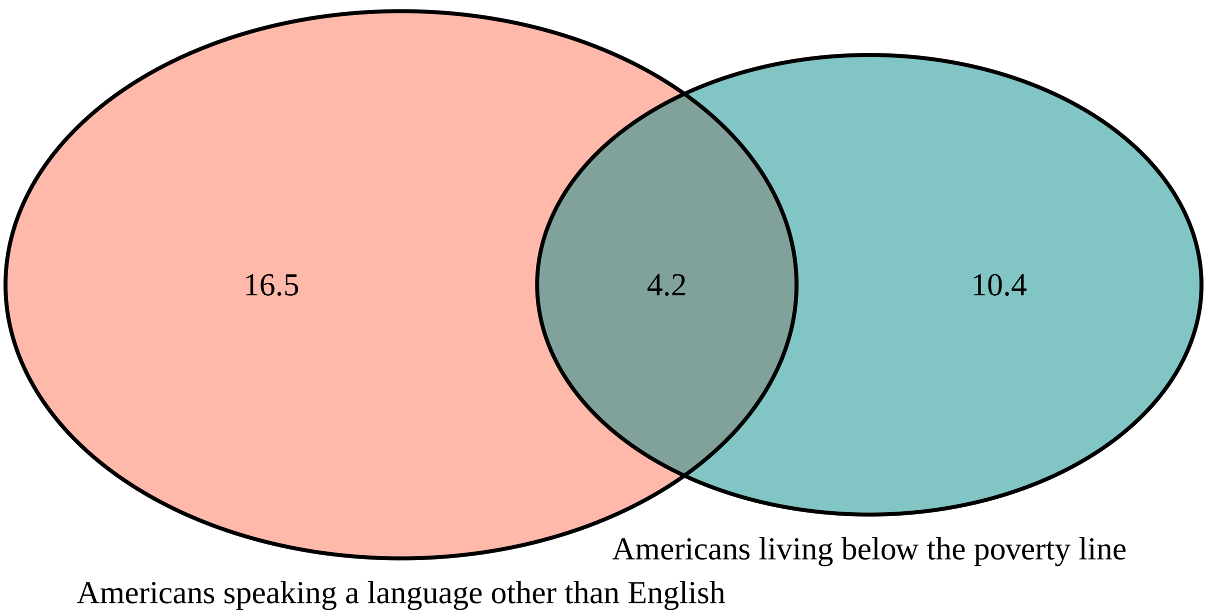
2.8 Poverty and language. The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English (foreign language) at home, and 4.2% fall into both categories.

(a) Are living below the poverty line and speaking a foreign language at home disjoint?

No, the two outcomes are not disjoint because someone could have both outcomes. Living above the poverty line and living below the poverty line is disjointed because both cannot occur at the same time.

(b) Draw a Venn diagram summarizing the variables and their associated probabilities.

```r
library(VennDiagram)
draw.pairwise.venn(area1 = 14.6, area2 = 20.7, cross.area = 4.2,
                   category = c("Americans living below the poverty line",
                                "Americans speaking a language other than English"),
                   fill = c("cyan4", "coral1"), cat.pos = c(0,0))
```



(polygon[GRID.polygon.1], polygon[GRID.polygon.2], polygon[GRID.polygon.3], polygon[GRID.polygon.4], text[GRID.text.5], text[GRID.text.6], text[GRID.text.7], text[GRID.text.8], text[GRID.text.9])

(c) What percent of Americans live below the poverty line and only speak English at home?

```r
library(scales)
below_poverty <- .146
other_than_english <- .207
both <- .042

paste("The percent of Americans living below the poverty line",
      "and only speaking English is",
      percent(below_poverty - both))
```

[1] "The percent of Americans living below the poverty line and only speaking English is 10.4%"

(d) What percent of Americans live below the poverty line or speak a foreign language at home?

```r
paste("The percent of Americans living below the poverty line",
      "or speaking a foreign language at home is",
      percent(below_poverty + other_than_english - both))
```

[1] "The percent of Americans living below the poverty line or speaking a foreign language at home is 31.1%"

(e) What percent of Americans live above the poverty line and only speak English at home?

```r
paste("The percent of Americans living ABOVE the poverty line",
      "and only speaking English at home is",
      percent(1 - (below_poverty + other_than_english - both)))
```

[1] "The percent of Americans living ABOVE the poverty line and only speaking English at home is 68.9%"

(f) Is the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home?

Yes, the outcomes are independent because living below the poverty line does not provide useful information on if the person also speaks a foreign language. We can see that the two intersect but they are independent of each other.

2.20 Assortative mating. Assortative mating is a nonrandom mating pattern where individuals with similar genotypes and/or phenotypes mate with one another more frequently than what would be expected under a random mating pattern. Researchers studying this topic collected data on eye colors of 204 Scandinavian men and their female partners. The table below summarizes the results. For simplicity, we only include heterosexual relationships in this exercise.[65]

		Partner (female)			
		Blue	Brown	Green	Total
Self (male)	Blue	78	23	13	114
	Brown	19	23	12	54
	Green	11	9	16	36
	Total	108	55	41	204

```r
eyes <- c("Blue","Brown", "Green")
self_eyes <- paste("self", eyes)
partner_eyes <- paste("partner", eyes)
values <- c(78,23,13,19,23,12,11,9,16)
```

```r
M <- matrix(values, ncol = 3, nrow = 3, byrow = TRUE)
rownames(M) <- self_eyes
colnames(M) <- partner_eyes
M
```

```
##            partner Blue partner Brown partner Green
## self Blue            78            23            13
## self Brown           19            23            12
## self Green           11             9            16
```

(a) What is the probability that a randomly chosen male respondent or his partner has blue eyes?

```r
paste("The probability that a randomly chosen male respondent",
      "or his partner have blue eyes is",
      percent((sum(M[1,]) + sum(M[,1]) - M[1,1]) / (sum(M[1,], M[2,],M[3,]))))
```

[1] "The probability that a randomly chosen male respondent or his partner have blue eyes is 70.6%"

(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?

```r
paste("The probability that a randomly chosen male respondent",
      "and his partner have blue eyes is",
      percent(M[1,1]/sum(M[1,])))
```

[1] "The probability that a randomly chosen male respondent and his partner have blue eyes is 68.4%"

(c) What is the probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes? What about the probability of a randomly chosen male respondent with green eyes having a partner with blue eyes?

```r
paste("The probability that a randomly chosen male respondent",
      "with brown eyes has a partner with blue eyes is",
      percent((M[2,1])/sum(M[2,])))
```

[1] "The probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes is 35.2%"

```r
paste("The probability that a randomly chosen male respondent",
      "with green eyes has a partner with blue eyes is",
      percent(M[3,1]/sum(M[3,])))
```

[1] "The probability that a randomly chosen male respondent with green eyes has a partner with blue eyes is 30.6%"

(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

It does not appear that the eye colors of male respondents and their partners are independent, this is evident in the higher probabilities that partners will have the same eye color. Essentially, the eye color of the male respondent provides some information regarding the eye color of the partner, which is that they are same, and that means the eye colors are not independent.

2.30 Books on a bookshelf. The table below shows the distribution of books on a bookcase based on whether they are nonfiction or fiction and hardcover or paperback.

		Format		
		Hardcover	Paperback	Total
Type	Fiction	13	59	72
	Nonfiction	15	8	23
	Total	28	67	95

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.

P(hardcover) 28 / 95 = 0.294

P(paperback fiction) 59 / 94 = 0.627

```
paste("The probability of drawing a hardcover book first then a",
      "paperback fiction book second when drawing without replacement is",
      percent((28 / 95) * (59/94) ))
```

[1] "The probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement is 18.5%"

(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.

```
P(Hardcover fiction & Hardcover)
Hardcover fiction
13 / 95
= 0.1368421

Hardcover
27 / 94
= 0.287234

P(Hardcover Fiction & Hardcover) = 0.0393057

P(Paperback fiction & Hardcover)
Paperback Fiction
59 / 95
= 0.6210526

Hardcover
28 / 94
=0.2978723

P(Paperback fiction & Hardcover) = 0.1849944
```

```
paste("The probability ofdrawing a fiction book first",
      "and then a hardcover book second,",
      "when drawing without replacement is",
      percent(((13 / 95) * (27 / 94)) + ((59/95) * (28 / 94))))
```

[1] "The probability ofdrawing a fiction book first and then a hardcover book second, when drawing without replacement is 22.4%"

(c) Calculate the probability of the scenario in part (b), except this time complete the calculations under the scenario where the first book is placed back on the bookcase before randomly drawing the second book.

```
P(Fiction)
72 / 95 = 0.757

P(Hardcover)
28 / 95 = 0.294

.757 * .294 = 0.223
22.3%
```

(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case

The reason the answers are so close is due to the number of books involved. We are only selecting a small number of books (2) and the hardcover books represent a smaller population overall.

2.38 Baggage fees. An airline charges the following baggage fees: \$25 for the first bag and \$35 for the second. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage and 12% have two pieces. We suppose a negligible portion of people check more than two bags.

(a) Build a probability model, compute the average revenue per passenger, and compute the corresponding standard deviation.

Event	X	P(X)	X - P(X)	(X - E(X))2	(X - E(X))2 * P(X)
No luggage	\$0.00	.54	0	(0 - \$15.70)2 = 246.49	246.49 * .54 = 133.1046
One Bag	\$25.00	.34	8.50	(25 - 15.70)2 = 86.49	86.49 * .34 = 29.406
Two Bags	\$60.00	.12	7.20	(60 - 15.70)2 = 1962.49	1962.49 * .12 = 235.4988
			E(X) - \$15.70		**V(X) - 398.0094**
					SD(X) = sqrt(V(X)) = 19.95017

Average Revenue per Passenger = \$15.70 Standard Deviation = 19.95017

(b) About how much revenue should the airline expect for a flight of 120 passengers? With what standard deviation? Note any assumptions you make and if you think they are justified

The expected revenue for a flight of 120 passengers is 1884.

The standard deviation would remain the same at 19.95017.

We are assuming that this is an appropriate probability to use for this flight. The flight could be going to a vacation destination in which case we might have more two bag passengers that expected because the population characteristics has variability unaccounted for in our model.

2.44 Income and gender. The relative frequency table below displays the distribution of annual total personal income (in 2009 inflation-adjusted dollars) for a representative sample of 96,420,486 Americans. These data come from the American Community Survey for 2005-2009. This sample is comprised of 59% males and 41% females.

Income	*Total*
$1 to $9,999 or loss	2.2%
$10,000 to $14,999	4.7%
$15,000 to $24,999	15.8%
$25,000 to $34,999	18.3%
$35,000 to $49,999	21.2%
$50,000 to $64,999	13.9%
$65,000 to $74,999	5.8%
$75,000 to $99,999	8.4%
$100,000 or more	9.7%

(a) Describe the distribution of total personal income.

The distribution appears to be normally distributed (which is unusual since income is never normally distributed), with a min at $1 and a max of $100,000 or more. The largest population is located at the center bin of $35,000 to $49,999 income. The increase on the right side may be due to the inclusion of outliers beyond $100,000 and the addition of more bins might be able to provide a smoother and better distributed curve.

(b) What is the probability that a randomly chosen US resident makes less than $50,000 per year?

Income	Total
$1 to $9,999 or loss	2.2%
$10,000 to $14,999	4.7%
$15,000 to $24,999	15.8%
$25,000 to $34,999	18.3%
$35,000 to $49,999	21.2%
Sum = 0.622	

```
P(x) = 62.2%
```

(c) What is the probability that a randomly chosen US resident makes less than $50,000 per year and is female? Note any assumptions you make.

Income	Total
$1 to $9,999 or loss	2.2%
$10,000 to $14,999	4.7%
$15,000 to $24,999	15.8%
$25,000 to $34,999	18.3%
$35,000 to $49,999	21.2%
Sum = 0.622	

```
P(x) = 62%
Males = 59% = .62 * .59 = 0.36698 = 37%
Females = 41% = .62 * .41 = 0.25502 = 25%

Making less than 50k & female = 25%
```

```
Assuming that this information is normally distributed for females and males.
```

(d) The same data source indicates that 71.8% of females make less than $50,000 per year. Use this value to determine whether or not the assumption you made in part (c) is valid.

My assumption is not valid, if we looked at the distribution for females only, the curve would have a right skew and we could not draw the same conclusion that I did in the previous answer. While personal income is normally distributed amoung US residents for this question, it is not normally distributed by sex. Also, there may be a sampling bias and there could be the possiblity that more poor females were surveyed than one would expect in the actual population.